

Defining optimal sampling effort for large-scale monitoring of invasive alien plants: a Bayesian method for estimating abundance and distribution

Cang Hui^{1*}, Llewellyn C. Foxcroft^{1,2}, David M. Richardson¹ and Sandra MacFadyen²

¹Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa; and ²Scientific Services, South African National Parks, Private Bag X402, Skukuza 1350, South Africa

Summary

1. Monitoring the abundance and spatial structure of invasive alien plant populations is important for designing and measuring the efficacy of long-term management strategies. However, methods for monitoring over large areas with minimum sampling effort, but with sufficient accuracy, are lacking. Although sophisticated sampling techniques are available for increasing sampling efficiency, they are often difficult to implement for large-scale monitoring, thus necessitating a robust yet practical method.

2. We explored this problem over a large area (*c.* 20 000 km²), using ad hoc presence–absence records routinely collected over 4 years in Kruger National Park (KNP), South Africa. Using a Bayesian method designed to solve the pseudo-absence (or false-negative) dilemma, we estimated the abundance and spatial structure of all invasive alien plants in KNP. Five sampling schemes, with different spatially weighted sampling efforts, were assessed and the optimal sampling effort estimated.

3. Although most taxa have very few records (50% of the species have only one record), the more abundant species showed a log-normal species-abundance distribution, with the 29 most abundant taxa being represented by an estimated total of 2.22 million individuals, with most exhibiting positive spatial autocorrelation.

4. Estimations from all sampling schemes approached the real situation with increasing sampling effort. An equal-weighted (uniform) sampling scheme performed best for abundance estimation (optimal efforts of 68 records per km²), but showed no advantage in detecting spatial autocorrelation (247 records per km² required). With increasing sampling effort, the accuracy of abundance estimation followed an exponential form, whereas the accuracy of distribution estimation showed diverse forms. Overall, a power law relationship between taxon density (as well as the spatial autocorrelation) and the optimal sampling effort was determined.

5. *Synthesis and applications.* The use of Bayesian methods to estimate optimal sampling effort indicates that for large-scale monitoring, reliable and accurate schemes are feasible. These methods can be used to determine optimal schemes in areas of different sizes and situations. In a large area like KNP, the uniform equal-weighted sampling scheme performs optimally for monitoring abundance and distribution of invasive alien plants, and is recommended as a protocol for large-scale monitoring in other protected areas as well.

Key-words: abundance estimation, Bayesian estimation, biological invasions, invasive alien plant, large-scale monitoring, protected areas, pseudo-absence, spatial autocorrelation

Introduction

Protected areas are an important component of global biodiversity conservation strategies (Gaston *et al.* 2008). Biological invasions are a major direct driver of biodiversity loss,

changes in ecosystem services (Wilcove *et al.* 1998) and biotic homogenization (Brook, Sodhi & Bradshaw 2008), and threaten the ecological integrity of most protected areas. Managing those invasive species that cause negative impacts is one of the major tasks of managers in many protected areas. Effective management of invasive species demands multi-layer

*Correspondence author. E-mail: chui@sun.ac.za

initiatives linked with research and monitoring. For example, the management of invasive alien plants (IAPs) in South Africa's Kruger National Park (KNP) has been supported by research in the following areas: detailed studies on the determinants and dynamics of spread of key taxa (Foxcroft *et al.* 2004; Foxcroft & Rejmánek 2007; Foxcroft, Richardson & Wilson 2008), assessment of the risks posed by alien plants in watersheds upstream of the park (Foxcroft, Rouget & Richardson 2007) and the effectiveness of the park boundary as a filter for invasive species (Foxcroft *et al.* 2011), and examination of the importance of issues pertaining to spatial scale in designing management plans (Foxcroft *et al.* 2009). Considerable efforts have also been made to initiate objective evaluation of all management initiatives to adapt and improve interventions (Biggs & Rogers 2003). A major challenge is to develop a cost-efficient monitoring strategy for this very large protected area, i.e. one that uses available records for reliable inference of the abundance and distributional structure of IAPs. The same challenge exists for protected areas around the world. This paper deals with fundamental requirements for an effective monitoring programme for IAPs. Although drawing on empirical data from KNP, the methodology can be applied in the general ecological research for estimating species abundance and distributions, and the results recommended for IAP management and control at regional scales.

Large-scale monitoring programmes (LSMPs) have obvious advantages over local-scale studies. First, sources of invasions can be included and identified in LSMPs, which is often impossible for local-scale studies. Secondly, spatial stochasticity is often minimized to facilitate the recognition of trends and patterns. Finally, environmental complexity and heterogeneity are often included to allow robust inference and projection that are compatible with policy making (Johnson 1993; Urquhart, Paulsen & Larsen 1998). Nevertheless, the implementation of LSMPs is often constrained by cost (e.g. Bottrill *et al.* 2009). Consequently, designing cost-efficient methods of inference and sampling schemes (protocols), that are accurate enough to inform various management and planning activities, is crucial for improving the effectiveness of LSMP. An efficient sampling scheme that requires the smallest sampling effort but which derives adequate monitoring results is urgently needed.

For assessing the impact of IAPs, the two most important variables that should be effectively estimated from such monitoring programmes are spatial distribution and abundance (Parker *et al.* 1999). Abundance is one of the most important measures of species conservation status (World Conservation Union 2001) and is a surrogate for ecological functioning (Gaston & Blackburn 2000; McGill *et al.* 2007). The spatial distribution of species reflects the dispersal processes and pathways of biological invasions, and is also a strong predictor of extinction risk and range contraction (Gaston & Fuller 2009). For instance, prioritization of areas for management using multi-criteria decision models often requires a reliable density map (and also maps of topography, disturbances and clearing history) as important input data to allow robust inference (Roura-Pascual *et al.* 2010). Better understanding of species abundance and distribution patterns can improve the monitor-

ing of changes in biodiversity (Strayer 1999; Yoccoz, Nichols & Boulinier 2001; Wilson *et al.* 2004), optimize conservation and control efficiency (Van Kleunen & Richardson 2007), and enable the assessment of impact and potential distribution of invasions (Parker *et al.* 1999).

The effectiveness of LSMPs can be improved by the careful design of sampling strategies (Carlson & Schmiegelow 2002; Thompson 2002; Fortin & Dale 2005). Issues that must be considered include the determination of an appropriate sampling effort, extent, unit size (grain) and sampling strategy (scheme, or spatial layout of the sampling unit); all these factors affect the potential of the initiative to detect different spatial patterns (Dungan *et al.* 2002; Hui, Veldtman & McGeoch 2010). For instance, adaptive cluster sampling enhances the efficiency compared to simple random sampling for estimating population densities of aggregated species (Thompson 1990). As a rule of thumb, sampling designs should be guided by the minimum requirement for adequate spatial analysis and inference (Fortin & Dale 2005), an issue which is likely to be difficult in a LSMP (Goodman 2003). As a commonly used and cost-efficient format for data recording (Brotons *et al.* 2004; Joseph *et al.* 2006), presence-absence data and ad hoc records have been adopted as the standard format in many monitoring programmes. In such cases, the choice of extent and grain is not an issue. Instead, we face the problem of false-negatives and pseudo-absences in the model inference (e.g. MacKenzie *et al.* 2002; Royle, Nichols & Kéry 2005). An efficient sampling design thus requires an adequate sampling effort for an efficient and practical sampling scheme that can provide enough information for accurate inference of abundance and the spatial distribution of IAPs.

The monitoring programme for IAPs in South Africa's KNP was chosen as a case study. This monitoring programme uses a system called CyberTracker that allows field rangers to report the location of IAPs and other features of interest (which count as "absence" records for IAPs) encountered on routine patrols (Foxcroft *et al.* 2009). We use the data gathered in this programme to determine the minimum sampling effort for monitoring IAPs in KNP. First, we designed a Bayesian model and estimated the abundance and spatial autocorrelation of IAPs by subdividing the landscape into lattices at a resolution of 4×4 km grid cells. Secondly, by assuming that the estimates from all records reflect real abundance and its distribution, a re-sampling simulation was designed for a variety of sampling schemes to evaluate their performance in model inference. Finally, the relationship between abundance estimation (also spatial autocorrelation) and sampling effort was identified to ensure we achieved satisfactory accuracy. The optimal sampling effort was presented as a function of the density and spatial correlation of the focal IAP species.

Materials and methods

STUDY AREA AND DATA

The KNP ($30^{\circ}53' - 32^{\circ}01'E$, $22^{\circ}19' - 25^{\circ}31'S$), stretching across $19\,485\text{ km}^2$, is one of the largest protected areas in the world that is

actively managed primarily for biodiversity conservation. More than 370 alien plant species have been recorded in KNP (Foxcroft *et al.* 2003; Foxcroft, Richardson & Wilson 2008). The goals of IAP management are to detect new invasions at an early stage and to respond rapidly to these, to maintain current invasions at low abundances where eradication is not feasible, and to undertake various actions to prevent further invasions (Foxcroft & Richardson 2003; Foxcroft & Freitag-Ronaldson 2007). The CyberTracker system (<http://www.cybertracker.org>), a user-friendly interface for PalmOS computers linked to geographical positioning systems (GPS), was introduced to facilitate the achievement of management goals in 2003, by providing rangers with a tool for management, and also high-precision data for research (MacFadyen 2005). In the KNP programme, this includes capturing information on animals, plants, water holes, poaching activity, fence condition, fire scars and numerous other features, including invasive alien plants. During their patrols, rangers move through the park in a haphazard fashion, recording the features they observe. The CyberTracker system is also set to take GPS readings at pre-defined time intervals. When recording a specific observation, other features, if present, will also be captured. Thus, timed points and records of other features can be used as 'absence' point data when analyzing alien plant invasions (see discussion in Foxcroft *et al.* 2009). We used data from 2004 to 2007 which comprises 2 360 419 records (including 27 777 presences and 2 332 642 absences) (Fig. 1a,b). One hundred and sixty-seven IAP species were included in the presence records; most records were for *Opuntia stricta* (72.1%) and *Lantana camara* (8.4%), and 119 species have fewer than 10 records. The extremely low occurrence (1.2%) of IAPs does not indicate insufficient sampling intensity or sampling preference (given the vast amount of records and the sampling protocol for rangers), but can be, in part, ascribed to the current policy of preventing and managing alien plant invasions. However, invasions of many species are currently in an early phase, with the potential for rapid expansion.

BAYESIAN ESTIMATION

After dividing the KNP into grids, we transferred the presence-absence records for each species into the detection-nondetection grids and faced the problem of pseudo-absence data: cells with only

absence records are not necessarily true absence of the focal species (e.g. MacKenzie *et al.* 2002, 2003; Tyre *et al.* 2003; Royle, Nichols & Kéry 2005; Pearce & Boyce 2006). Although methods are available for abundance inference using true presence-absence binary data (e.g. Nachman 1981; Wright 1991; He & Gaston 2003; Hui, McGeoch & Warren 2006; Hui *et al.* 2009; Borregaard & Rahbek 2010), this pseudo-absence dilemma still requires serious consideration. Solutions to this dilemma can arise from two statistical philosophies (Anderson 2008): maximum likelihood and Bayesian methods. The maximum likelihood method combines two binomial processes in forming a joint likelihood for the probability of absence and detection for each cell. This can then be estimated by maximizing the logarithmic likelihood by assuming a constant point-detection rate (MacKenzie *et al.* 2002, 2003) or a pre-defined distribution of species abundance (Royle & Nichols 2003; Zhou & Griffiths 2007). This is an appropriate approach for small-scale studies in homogeneous landscapes, but not for LSMPs because of the inherent spatial heterogeneity and some technical drawbacks in the abundance estimation (Warren, McGeoch & Chown 2003; Conlisk, Conlisk & Harte 2007). In contrast, the Bayesian method has also been applied to estimating avian abundance and occurrence in repeated surveys with encounter histories for each cell (Royle *et al.* 2007; Royle & Dorazio 2008). We thus followed this statistical philosophy and presented a Bayesian method to the pseudo-absence dilemma and abundance estimation for each IAP species in the KNP.

For a cell of size a with a number of n reported records (including x presences), if each record only corresponds to one non-repetitive expeditious visit of an area of δ with perfect detection, then a number of $M = a/\delta$ records are needed for obtaining the full information of this cell. This is reasonable assumption given that rangers are pre-trained for identifying specific IAPs and have to move rapidly across vast areas and that the occurrence of particular IAP species in the records is much lower than 1.2%. In the calculation, we set $a = 4 \times 4$ km (with a total of 1333 cells in the KNP) and M the maximum number of records within cells for practical reasons (i.e. $M = 72\,203$, indicating a 100% detection rate when rangers report a record within 8.4 m radius; less than 1% cells with more than 10 000 records). We assume that there are actually N presences (i.e. the true underlying abundance) once we had the full-information of the M

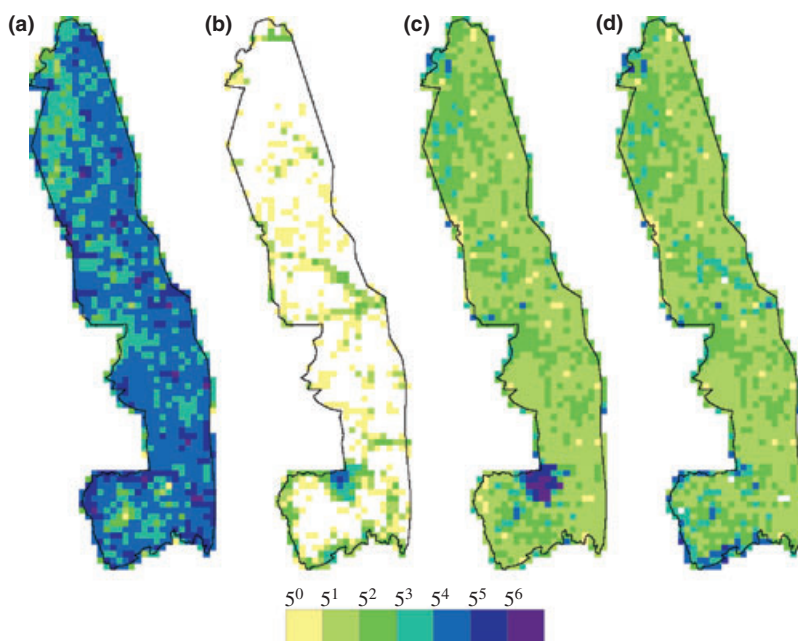


Fig. 1. The number of records (a) and presence records (b) in Kruger National Park, as from the CyberTracker Programme, plotted at a resolution of 4×4 km grids. The spatial structure of the number of *Opuntia stricta* (c) and *Lantana camara* (d), as predicted from the Bayesian model. The spatial structures of other invasive species are presented in Fig. S2 Supporting Information.

records. The detection rate of one random sampling (record) in this cell thus equals N/M (i.e. the detection rate reflects the true abundance of a species in the cell; MacKenzie *et al.* 2005), which often differs from the occurrence of presence in samples, x/n . Clearly, the probability of finding x presences in the n reported records, knowing that there are N presences in the full-information M records, follows a hyper-geometric distribution: $\text{prob}(x|N) = C_x^N C_{n-x}^{M-N} / C_n^M$ (where C_x^N is the binomial coefficient, $N!/(x!(N-x)!)$, for ' N choose x '; here we treated the in-cell number of sampling records n and M as two known parameters and thus deleted in the left-side probability notation for brevity). In doing so, we neglect the in-cell heterogeneity (note that the between-cell spatial heterogeneity remains the same as in the cell-specific detection rate) and assume M the carrying capacity for all IAPs species. Such relatively reasonable compromises enable the possibility for a species-specific inference. Therefore, the probability distribution for in-cell abundance N given that x presences have been reported in the n samples can be estimated by the Bayesian rule:

$$\text{prob}(N|x) = \frac{\text{prob}(x|N) \cdot \text{prob}(\cdot|N)}{\sum_{y=x}^{M-n+x} \text{prob}(\cdot|y) \cdot \text{prob}(x|y)}, \quad \text{eqn 1}$$

where $\text{prob}(\cdot|N)$ is a prior probability of N presences in the cell regardless of any sampling information. In the calculation, we set this prior probability as a combination of the uninformative prior (Jaynes 1968) and the Poisson model (for alleviating the zero-inflation problem; Royle, Nichols & Kéry 2005): $\text{prob}(\cdot|N) = (1 - \exp(-d \cdot M)) / (N \cdot \sum_{y=1}^M 1/y)$ for $N \geq 1$, and $\text{prob}(\cdot|N) = \exp(-d \cdot M)$ for $N = 0$, where d stands for the occurrence of a certain species (that is, the proportion of presences records for a focal species in all the records). For the pseudo-absence dilemma, the probability of absence in the cell, given that all n records reported were absence, is (see Fig. S1 in Supporting Information for an illustration):

$$\text{prob}(N=0|x=0) = \frac{\exp(-d \cdot M)}{\sum_{y=0}^{M-n} \text{prob}(\cdot|y) \frac{C_{n-y}^{M-y}}{C_n^M}}, \quad \text{eqn 2}$$

The mean and variance of the abundance estimated in the cell can thus be given by $E(N) = \sum N \cdot \text{prob}(N|x)$ and $V(N) = \sum N^2 \text{prob}(N|x) - E(N)^2$, respectively. To estimate the abundance and its variation for focal species in the entire KNP, we did not use the unbiased Horvitz–Thompson estimator as all cells in KNP have been sampled (Thompson 2002) and also that the above in-cell estimation has considered sampling intensity within each cell, which resembles stratified sampling. Consequently, the expected abundance of the focal species in the KNP is the sum of all the in-cell mean abundance, $E(N)$, across the entire KNP. Furthermore, by assuming the independence of the number of individuals among different cells (i.e. setting the covariance to zero), we can estimate the minimum variance of total abundance as the sum of the in-cell variances across the landscape, from which the confidence interval of total abundance (and density) can be derived.

The above process yielded a spatially heterogeneous map of in-cell abundance for each species at a resolution of 4×4 km. This spatial heterogeneity of cell-specific local abundance and detection rate reflects spreading history, demographics and the spatial structure of suitable habitat for focal species (Guisan & Thuiller 2005). To describe the spatial heterogeneity of species distributions, we calculated the spatial autocorrelation of focal species using the first-distance class Moran's (1950) I index that describes the degree to which points in space are correlated (Hui, Veldtman & McGeoch

2010). The mean and variance of Moran's I can be estimated by permutation tests (e.g. Rodríguez & Delibes 2002). Specifically, the mean (i.e. equals the negative reciprocal of the number of cells minus one) and variance of spatial randomness are only dependent on the spatial configuration of the landscapes, not on the number of individuals within cells. The Moran's I for each species was then calculated using the mean abundance within cells ($E(N)$).

OPTIMAL SAMPLING EFFORT

Since incorporating the detection rate in abundance estimation for LSMPs is not common, choosing appropriate sampling schemes and optimal efforts for the above Bayesian estimation model, and other similar models, is necessary (e.g. Thompson & Seber 1996). The following re-sampling tests take the abundance estimated from the above Bayesian method as true and thus enable the accuracy of estimates under different sampling efforts to be evaluated (see Appendix S1 in Supporting Information for tests using simulated species). To determine the most-efficient sampling scheme, we first let s denote the total number of records reported (a sum of presences and absences), i.e. the sampling effort. For each unit of sampling effort, only one cell can be visited and the chance of reporting a presence record is equal to the detection rate, i.e. the proportion of expected true presences estimated for the focal species within the cell ($E(N)/M$). Five sampling schemes were initially examined, including weighted (the sampling effort is allocated according to observations from the CyberTracker data), uniform (i.e. equal-weighted cells), additive (the ranger tends to visit the cells having more presence records), elusive (the ranger will try to avoid visiting the cells with presence records) and random-walk (the ranger will randomly choose a cell adjacent to the cell visited at the last time). Although other sophisticated sampling schemes exist with which to increase the sampling efficiency (e.g. adaptive cluster sampling; Dryver & Thompson 2005), they are difficult to implement in practice, especially given the species-specific spatial variation and the multifaceted targets of KNP's monitoring programme. The abundance and spatial autocorrelation of IAPs were estimated for different sampling effort using the Bayesian method. The performance of such re-sampling was evaluated by the accuracy index, defined as the proportion deviation in estimating the abundance and spatial autocorrelation from those estimations using full records: $A = \text{Abs}[(\text{Estimate}_{\text{sample}} - \text{Estimate}_{\text{full}}) / \text{Estimate}_{\text{sample}}]$ (Hui, McGeoch & Warren 2006). The optimal sampling effort was thus defined as the minimum sampling effort for estimating abundance and distributional structure at a satisfactory level (we chose $A < 0.05$). Of more importance though, we examined how the estimations change as a function of sampling effort.

Results

The abundance and spatial autocorrelation of all 167 IAP species were calculated by the Bayesian method at a resolution of 4×4 km. The data were dominated by IAP species with only a single presence record and density estimated less than 0.1 km^{-2} (grey bar in Fig. 2). The abundance estimations of the rest of the IAP species formed a log-normal form of species-abundance distribution (white bars in Fig. 2; Kolmogorov–Smirnov test, $D = 0.1$, $P > 0.05$). We only present data for those species with at least 22 reported presence records, representing a relatively reliable estimation for the 29 most recorded species (Table 1). These 29 species represent 95% of the presence records, with a total of 2.22 million IAP individuals estimated. The mean density of these 29 species ranges from

1.29 km⁻² for *Cardiospermum grandiflorum* to 28.2 km⁻² for *O. stricta*. Although not significant, six species were found to have a negative value of Moran's *I* (Table 1; note the expected

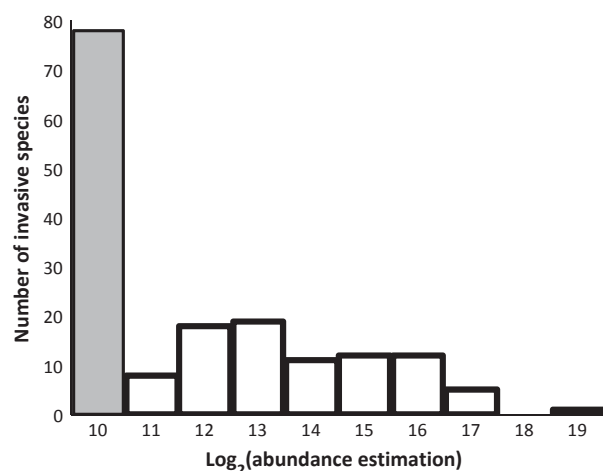


Fig. 2. Species-abundance distribution of the 167 invasive plant species in the Kruger National Park, as estimated from the Bayesian model. The grey bar shows an overwhelming number of rare invasive plant species; by excluding the rare species the white bars follow a log-normal shape.

value for randomness is $I = -0.00075$); other species were significantly positively autocorrelated, including the two most abundant species: *O. stricta* and *L. camara* (Fig. 1c,d; see Fig. S2 for others).

Using randomization tests, we showed that the Bayesian estimations for all five sampling schemes approached the known abundance and distribution for simulated species, with the uniform sampling scheme performing best (Appendix S1). Therefore, we only demonstrated the results from the uniform and weighted sampling schemes in detail; these schemes represent an ideal and a self-organized realistic sampling scheme respectively. Both schemes can also be easily implemented in future monitoring (i.e. equal weight or maintaining current patrol protocol). With the increase in sampling effort, estimations of these 29 species from both uniform and weighted sampling schemes, as depicted by the abundance–rank curves (Fig. 3a) and species–aggregation curves (Fig. 3b), converged to the original Bayesian estimations, yet with different converging rates and directions. Although both schemes tend to overestimate species abundance at low sampling effort, estimations from the weighted sampling scheme converged at a much slower rate than with uniform sampling, with 0.5 (2) million records of weighted sampling producing almost the same results as 0.1 (0.5) million records of uniform sampling (Fig. 3a). This suggests that equal-weight uniform sampling is

Table 1. The number of records, abundance (with the 95% confidence interval of density, km⁻²) and distribution estimations (Moran's *I*; all positive value are significantly autocorrelated), as predicted from the Bayesian model, and the optimal sampling effort for abundance estimation (OSEA; records per km²) and distribution estimation (OSED; records per km²), as predicted from the relationship between sampling effort and these estimations (Figs 4, S3 and S4), for the top 29 recorded invasive species in Kruger National Park

| Species | Records | Abundance | 95% CI | 95% CI | Moran's <i>I</i> | OSEA | OSED |
|-----------------------------------|---------|-----------|--------|--------|------------------|--------|--------|
| <i>Opuntia stricta</i> | 20029 | 549243 | 25.877 | 30.499 | 0.0638 | 21.59 | 3.0 |
| <i>Lantana camara</i> | 2332 | 144885 | 5.569 | 9.303 | 0.0356 | 35.60 | 24.7 |
| <i>Opuntia</i> spp. | 1483 | 114205 | 4.019 | 7.703 | 0.0284 | 41.39 | 23.3 |
| <i>Chromolaena odorata</i> | 327 | 80234 | 2.271 | 5.965 | -0.0007 | 48.66 | 81.1 |
| <i>Pistia stratiotes</i> | 218 | 76485 | 2.089 | 5.762 | 0.0079 | 51.81 | 71.1 |
| <i>Parthenium hysterophorus</i> | 204 | 72736 | 1.903 | 5.563 | -0.0006 | 52.63 | 114.4 |
| <i>Eichhornia crassipes</i> | 199 | 78291 | 2.181 | 5.855 | 0.0111 | 51.76 | 64.0 |
| <i>Xanthium strumarium</i> | 176 | 72806 | 1.907 | 5.566 | -0.0008 | 51.36 | 109.8 |
| <i>Azolla filiculoides</i> | 160 | 70696 | 1.803 | 5.454 | 0.0053 | 53.41 | 91.9 |
| <i>Argemone</i> spp. | 135 | 69316 | 1.740 | 5.375 | 0.0049 | 53.73 | 137.8 |
| <i>Senna</i> spp. | 110 | 70842 | 1.830 | 5.442 | 0.0062 | 59.20 | 165.9 |
| <i>Xanthium</i> spp. | 98 | 65051 | 1.594 | 5.083 | 0.0052 | 61.90 | 193.4 |
| <i>Cardiospermum halicacabum</i> | 96 | 69566 | 1.775 | 5.366 | 0.0070 | 62.43 | 136.5 |
| <i>Argemone ochroleuca</i> | 93 | 62947 | 1.456 | 5.006 | -0.0007 | 60.15 | 116.4 |
| <i>Ageratum</i> spp. | 88 | 62375 | 1.436 | 4.966 | 0.0066 | 58.26 | 135.1 |
| <i>Ricinus communis</i> | 83 | 61069 | 1.380 | 4.888 | 0.0061 | 59.94 | 168.2 |
| <i>Argemone mexicana</i> | 71 | 55093 | 1.110 | 4.545 | -0.0007 | 67.37 | 124.3 |
| <i>Catharanthus roseus</i> | 63 | 51872 | 0.977 | 4.347 | 0.0052 | 77.29 | 239.1 |
| <i>Arundo donax</i> | 63 | 51160 | 0.941 | 4.310 | 0.0048 | 78.28 | 263.3 |
| <i>Datura innoxia</i> | 59 | 50525 | 0.928 | 4.258 | 0.0060 | 79.10 | 262.1 |
| <i>Datura stramonium</i> | 46 | 44443 | 0.699 | 3.862 | 0.0072 | 82.24 | 411.7 |
| <i>Ageratum conyzoides</i> | 45 | 42221 | 0.597 | 3.737 | 0.0058 | 84.00 | 437.3 |
| <i>Melia azedarach</i> | 37 | 36384 | 0.377 | 3.358 | 0.0048 | 88.67 | 321.1 |
| <i>Senna occidentalis</i> | 35 | 34991 | 0.329 | 3.263 | 0.0046 | 88.90 | 343.3 |
| <i>Datura ferox</i> | 26 | 27555 | 0.022 | 2.806 | -0.0008 | 91.79 | 51.3 |
| <i>Zinnia peruviana</i> | 25 | 26713 | 0.054 | 2.688 | 0.0046 | 97.86 | 628.9 |
| <i>Nicotiana glauca</i> | 25 | 26690 | 0.053 | 2.687 | 0.0035 | 97.79 | 1339.7 |
| <i>Psidium guajava</i> | 24 | 25363 | 0.003 | 2.600 | 0.0044 | 98.84 | 909.2 |
| <i>Cardiospermum grandiflorum</i> | 23 | 25156 | 0.012 | 2.570 | 0.0041 | 100.00 | 184.3 |

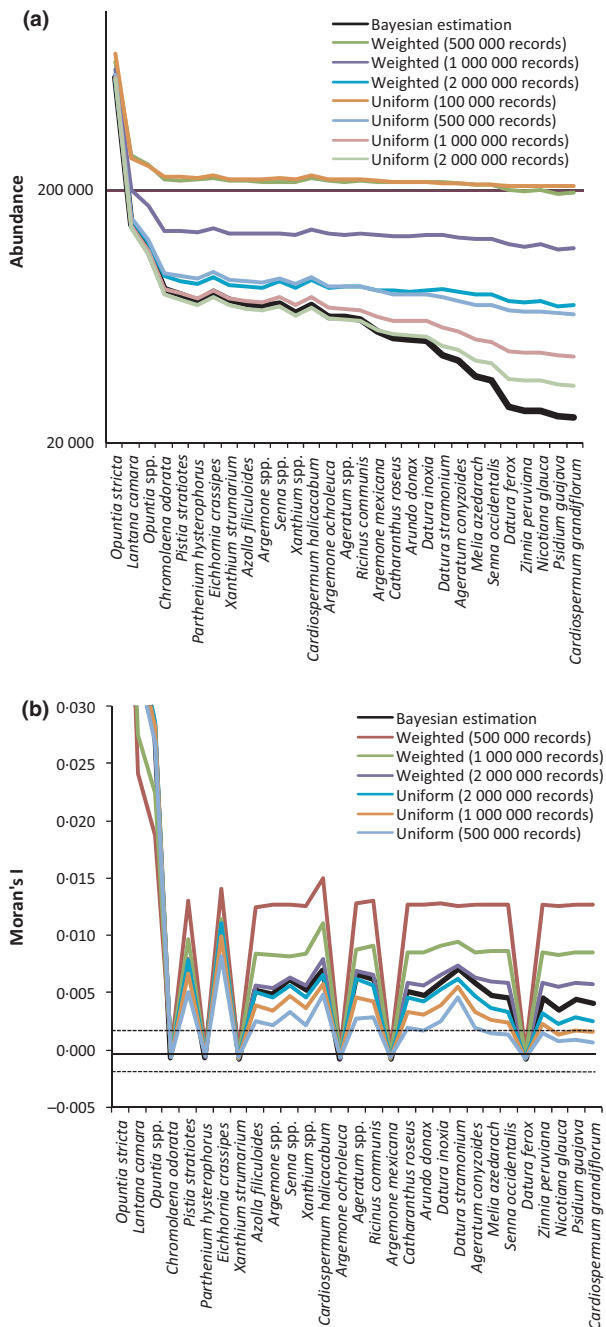


Fig. 3. The abundance–rank curves (a) and the species–aggregation curves (b) for the top-recorded 29 invasive plant species in Kruger National Park, under different sampling schemes and sampling efforts. The solid and dashed straight lines in (b) indicate the mean and confidence interval of Moran's I for randomly distributed species.

the most-efficient sampling scheme for abundance estimation. Furthermore, estimations from both sampling schemes performed well in estimating the spatial autocorrelation, although the weighted sampling scheme approached the original Bayesian estimation from the positive autocorrelation (or rather randomness) side and uniform sampling from negative autocorrelation side (Fig. 3b). Low sampling effort of uniform sampling tended to identify the spatial structure as random-

ness, in contrast to the tendency of significantly positive autocorrelation identified by low-effort weighted sampling (Fig. 3b).

The optimal sampling effort was calculated for the most-efficient sampling scheme, uniform (equal-weight) sampling, although similar results can be expected for weighted sampling. The relationship between sampling effort and the accuracy in abundance estimation followed a clear exponential relationship: $A = c_1 \text{Exp}(-c_2 \times s)$, where c_1 and c_2 are constants ($R^2 > 0.94$ for all 29 species; Fig. S3). The optimal sampling effort for abundance estimation (OSEA) was then estimated for $A = 0.05$, i.e. $\text{OSEA} = -\text{Ln}(0.05/c_1)/c_2$ (Table 1). A strong interspecific power law relationship between the OSEA and the IAP density emerged (Fig. 4a). The relationship between sampling effort and the accuracy in distribution estimation was much more diverse, from the unimodal form for all six species with negative values of Moran's I , to the power law and exponential forms for other IAP species (see Fig. S4 for details). The optimal sampling effort for distribution estimation (OSED) was then estimated accordingly for $A = 0.05$, which also led to a power law relationship between the OSED and IAP densities (Fig. 4b). A power law relationship between the OSEA and the OSED also emerged (Fig. 4c), suggesting an inner positive relationship between species density and aggregation. Overall, a spatially random, rare species required a much larger sampling effort to achieve the satisfactory level of accuracy. An average of 67.45 records per km² was required for abundance estimation of the top 29 species, 246.6 records per km² for distribution estimation (Table 1), comparing to the current level of 121.14 records per km².

Discussion

Large-scale monitoring programmes need cost-efficient methods for estimating the abundance and distribution of target species. Traditional mensuration methods of estimating abundance, such as mark–recapture techniques, are only useful at local scales (e.g. 0.1–10 km² for complete counts) due to the method of data collection and the associated costs. This has led to an increased interest in the use of binary (presence/absence) data for LSMPs (Brotons *et al.* 2004; Joseph *et al.* 2006). In this regard, two categories of abundance estimation models have been developed. The first is designed for true presence–absence binary data and includes the intraspecific occupancy–abundance relationship that is grounded in the ubiquitous positive correlation between species abundance and range size (Nachman 1984; Wright 1991; Gaston & Blackburn 2000; He & Gaston 2003; Hui & McGeoch 2007) and the scaling pattern of occupancy describes how adjacent occupied cells merge with increasing grain (Hartley & Kunin 2003; Hui, McGeoch & Warren 2006; Lennon *et al.* 2007; Gaston & Fuller 2009; Hui 2009). A multi-criteria test suggests the supremacy of the scaling-pattern-of-occupancy models over the occupancy–abundance–relationship models in estimating abundance and yielding macroecological patterns (Hui *et al.* 2009).

The other category is designed to tackle the pseudo-absence problem inherent in binary data and mainly uses the maximum

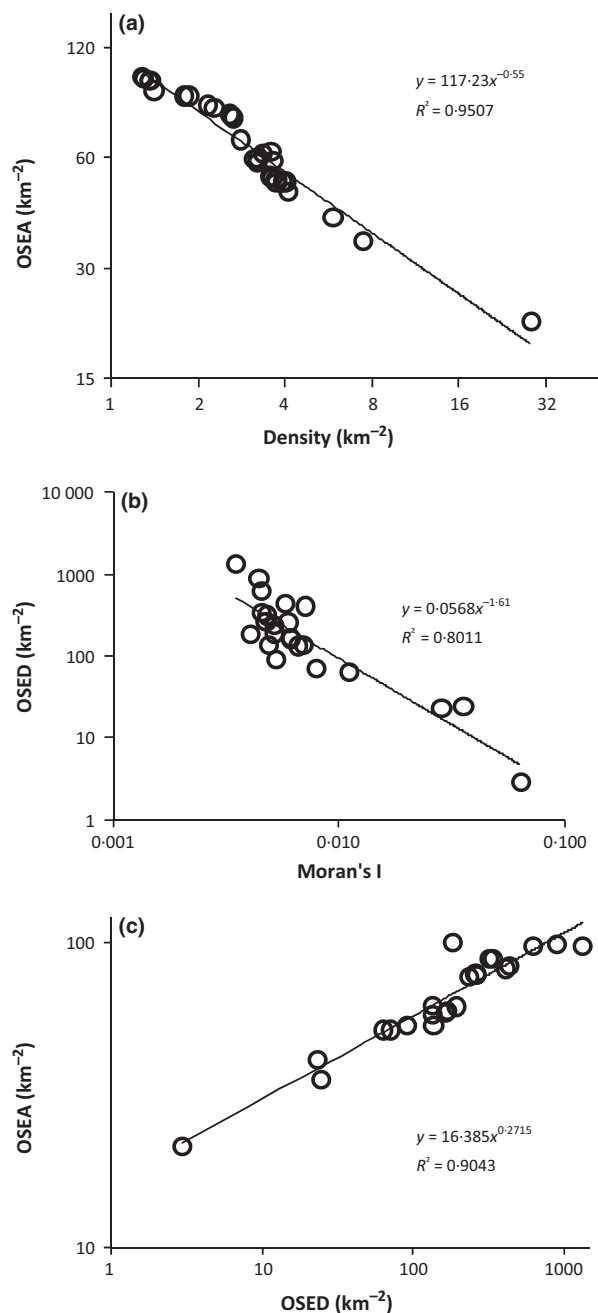


Fig. 4. The relationship between the density of invasive plants and the optimal sampling effort for abundance estimation (OSEA) (a), between the aggregation, as measured by Moran's I , and the optimal sampling effort for distribution estimation (OSER) (b), as well as between the OSEA and OSER (c).

likelihood method to reconcile the imperfect detection or presence-only data (e.g. Mackenzie *et al.* 2002; Royle, Nichols & Kéry 2005; Pearce & Boyce 2006; Zhou & Griffiths 2007; Nichols *et al.* 2008). Here, a compromise often has to be made for the alleviation of over-parameterization. The Bayesian model we present here contributes to the second category of models for dealing with the abundance estimation and pseudo-absence dilemma simultaneously, but follows the statistical philosophy of the Bayesian School. Such models need further investigation

given their obvious practical value (Royle *et al.* 2007; Royle & Dorazio 2008).

Studies on biological invasions at the level of entire community assemblages are limited (Mason & French 2008) due to the lack of evidence that invasive species can form a separate community in the invaded areas and also the difficulty of quantifying such a community if it exists. The species-abundance distribution (Fig. 2) in this study provides clues for both problems. Invasive species in KNP can be categorized into two groups: the recorded but not established (grey bar) and the recorded and established species (white bars in Fig. 2). This two-group bimodal species-abundance distribution is clearly different from the canonical unimodal log-normal (or zero-sum multinomial) shape that has been reported in numerous cases (since Preston 1948) and which is explained by various models (e.g. Gaston & Blackburn 2000; Hubbell 2001; Volkov *et al.* 2005; McGill *et al.* 2007; Haegeman & Etienne 2010).

Strong support for our two-group species-abundance distribution is provided by Magurran & Henderson (2003), who describe a very similar form for an estuarine fish community (Fig. 1 in Magurran & Henderson 2003). They explain such an excess of rare species by dividing the fish assemblage into two components: persistent and occasional species. Southwood (1996) reported a similar pattern in a Heteroptera insect community and divided the species into transient and core species, which is similar to Hanski's (1982) core-satellite hypothesis. Alien species move through a series of stages during the invasion process (Richardson *et al.* 2000) and often experience a time-lag before rapid expansion (Crooks 2005). This provides a natural break point in the species-abundance distribution to separate those established invaders from those that have been recorded but which are still in the time-lag. Magurran & Henderson (2003) further use the residence time (or persistence) to identify such a break point. Considering the importance of residence time in determining the potential range of an invading species (Wilson *et al.* 2007; Pyšek, Křivánek & Jarošík 2009), it is necessary to further investigate the relationship between residence time and IAP abundance, and verify whether this residence time can be used to categorize IAP species into core and satellite groups.

Ecological studies should always be carried out in an equal-weight fashion for further statistical inference (Quinn & Keough 2002). Our results confirmed this rule of thumb, even for LSMP. Weighted and sequential sampling schemes, such as additive and elusive, overestimated the abundance of IAPs (Appendix S1), and estimations from weighted sampling approached the real scenario at a much slower rate than the uniform sampling (Fig. 3). This suggests that the future patrol (or sampling) protocol in the CyberTracker monitoring programme (in terms of IAPs) should focus more on the under-sampled areas to counter the current highly skewed sampling weights (an average 121.14 records per km^2 , with a 95% confidence interval of 2.6 to 510.2 records per km^2). This means a minimum in-cell sampling effort of between 68 and 247 records per km^2 for monitoring abundance and distribution, respectively.

Tasked with conserving biological diversity, translating advances in science into management action is a key requisite and challenge for conservation agencies. Rapid advances in science are providing a wealth of insights on crucial facets of issues that are important to conservationists, yet few of these advances are operationalized for practical implementation in the field. One of the key challenges in conservation biology is finding appropriate ways of using information effectively (Richardson & Whittaker 2010). Data collected in the KNP for the purposes of both management and LSMPs provided the opportunity to explore options for improving the cost-efficiency of utilizing available data for monitoring and other purposes. The Bayesian method derived here provides management with an accurate and practical approach for monitoring invasive alien plants over a large area. This paves the way for improved prioritization of various interventions to address problems associated with biological invasions in protected areas.

Acknowledgements

We are grateful to G. Blanchet, G. Cruz-Piñón, F. He, K.J. Gaston, I. Kühn, W.E. Kunin and S. Hartley for comments and logistic help, and SANParks and the DST-NRF Centre of Excellence for Invasion Biology for financial support. C.H. acknowledges support from the NRF Blue Sky Programme. L.C.F. acknowledges support from the NFR Incentive Funds Programme. D.M.R. acknowledges support from the Hans Sigrist Foundation.

References

- Anderson, D.R. (2008) *Model Based Inference in the Life Science: A Primer on Evidence*. Springer, New York.
- Biggs, H.C. & Rogers, K.H. (2003) An adaptive system to link science, monitoring, and management in practice. *The Kruger Experience: Ecology and Management of Savanna Heterogeneity* (eds J. Du Toit, K.H. Rogers & H.C. Biggs), pp. 59–80. Island Press, Washington, DC, USA.
- Bottrill, M., Joseph, L.N., Carwardine, J., Bode, M.C., Cook, C., Game, E.T., Grantham, H., Kark, S., Linke, S., McDonald-Madden, E., Pressey, R.L., Walker, S., Wilson, K.A. & Possingham, H.P. (2009) Finite conservation funds mean triage is unavoidable. *Trends in Ecology and Evolution*, **24**, 183–184.
- Borregaard, M.K. & Rahbek, C. (2010) Causality of the Relationship between Geographic Distribution and Species Abundance. *Quarterly Review of Biology*, **85**, 3–25.
- Brook, B.B., Sodhi, N.S. & Bradshaw, C.J.A. (2008) Synergies among extinction drivers under global change. *Trends in Ecology and Evolution*, **23**, 453–460.
- Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H. (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, **27**, 437–448.
- Carlson, M. & Schmiegelow, F. (2002) Cost-effective sampling design applied to large-scale monitoring of Boreal Birds. *Conservation Ecology*, **6**, 11.
- Conlisk, E., Conlisk, J. & Harte, J. (2007) The impossibility of estimating a negative binomial clustering parameter from presence-absence data: A comment on He and Gaston. *American Naturalist*, **170**, 651–654.
- Crooks, J.A. (2005) Lag times and exotic species: the ecology and management of biological invasions in slow-motion. *Ecoscience*, **12**, 316–329.
- Dryver, A.L. & Thompson, S.K. (2005) Improved unbiased estimators in adaptive cluster sampling. *Journal of the Royal Statistical Society, Series B*, **67**, 157–166.
- Dungan, J.L., Perry, J.N., Dale, M.R.T., Legendre, P., Citron-Pousty, S., Fortin, M.-J., Jakomulska, A., Miriti, M. & Rosenberg, M.S. (2002) A balanced view of scale in spatial statistical analysis. *Ecography*, **25**, 626–640.
- Fortin, M.-J. & Dale, M.R.T. (2005) *Spatial Analysis: A Guide for Ecologists*. Cambridge University Press, Cambridge.
- Foxcroft, L.C. & Freitag-Ronaldson, S. (2007) Seven decades of institutional learning: managing alien plant invasions in the Kruger National Park, South Africa. *Oryx*, **41**, 160–167.
- Foxcroft, L.C., Jarošic, V., Pyšek, P., Richardson, D.M. & Rouget, M. (2011) Protected-area boundaries as filters of plant invasions. *Conservation Biology*, in press. Doi: 10.1111/j.1523-1739.2010.01617.x.
- Foxcroft, L.C. & Rejmánek, M. (2007) What helps *Opuntia stricta* invade Kruger National Park, South Africa: Baboons or elephants? *Applied Vegetation Science*, **10**, 265–270.
- Foxcroft, L.C. & Richardson, D.M. (2003) Managing alien plant invasions in the Kruger National Park, South Africa. *Plant Invasions: Ecological Threats and Management Solutions* (eds L. Child, J.H. Brock, G. Brundu, K. Prach, P. Pyšek, P.M. Wade & M. Williamson), pp. 385–404. Backhuys Publishers, Leiden, The Netherlands.
- Foxcroft, L.C., Richardson, D.M. & Wilson, J.R.U. (2008) Ornamental plants as invasive aliens: problems and solutions in the Kruger National Park, South Africa. *Environmental Management*, **41**, 32–51.
- Foxcroft, L.C., Rouget, M. & Richardson, D.M. (2007) Risk assessment of riparian plant invasions into protected areas. *Conservation Biology*, **21**, 412–421.
- Foxcroft, L.C., Henderson, L., Nichols, G.R. & Martin, B.W. (2003) A revised list of alien plants for the Kruger National Park. *Koedoe*, **46**, 21–44.
- Foxcroft, L.C., Rouget, M., Richardson, D.M. & MacFadyen, S. (2004) Reconstructing 50 years of *Opuntia stricta* invasion in the Kruger National Park, South Africa: environmental determinants and propagule pressure. *Diversity and Distributions*, **10**, 427–437.
- Foxcroft, L.C., Richardson, D.M., Rouget, M. & MacFadyen, S. (2009) Patterns of alien plant distribution at multiple spatial scales in a large national park: implications for ecology, management and monitoring. *Diversity and Distributions*, **15**, 367–378.
- Gaston, K.J. & Blackburn, T.M. (2000) *Pattern and Process in Macroecology*. Blackwell Science, Oxford.
- Gaston, K.J. & Fuller, R.A. (2009) The sizes of species' geographic ranges. *Journal of Applied Ecology*, **46**, 1–9.
- Gaston, K.J., Jackson, S.F., Cantú-Salazar, L. & Cruz-Piñón, G. (2008) The ecological performance of protected areas. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 93–113.
- Goodman, P.S. (2003) Assessing management effectiveness and setting priorities in protected areas in KwaZulu-Natal. *BioScience*, **53**, 843–850.
- Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.
- Haegeman, B. & Etienne, R.S. (2010) Self-consistent approach for neutral community models with speciation. *Physical Review E*, **81**, 031911.
- Hanski, I. (1982) Dynamics of regional distribution - the core and satellite species hypothesis. *Oikos*, **38**, 210–221.
- Hartley, S. & Kunin, W.E. (2003) Scale dependency of rarity, extinction risk, and conservation priority. *Conservation Biology*, **17**, 1559–1570.
- He, F. & Gaston, K.J. (2003) Occupancy, spatial variance, and the abundance of species. *American Naturalist*, **162**, 366–375.
- Hubbell, S.P. (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton.
- Hui, C. (2009) On the scaling pattern of species spatial distribution and association. *Journal of Theoretical Biology*, **261**, 481–487.
- Hui, C. & McGeoch, M.A. (2007) Capturing the “droopy-tail” in the occupancy-abundance relationship. *Ecoscience*, **14**, 103–108.
- Hui, C., McGeoch, M.A. & Warren, M. (2006) A spatially explicit approach to estimating species occupancy and spatial correlation. *Journal of Animal Ecology*, **75**, 140–147.
- Hui, C., Veldtman, R. & McGeoch, M.A. (2010) Measures, perceptions and scaling patterns of aggregated species distributions. *Ecography*, **33**, 95–102.
- Hui, C., McGeoch, M.A., Reyers, B., le Roux, P.C., Greve, M. & Chown, S.L. (2009) Extrapolating population size from the occupancy-abundance relationship and the scaling pattern of occupancy. *Ecological Applications*, **19**, 2038–2048.
- Jaynes, E.T. (1968) Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, **4**, 227–241.
- Johnson, S.P. (1993) *The Earth Summit: the United Nations Conference on Environment and Development*. Graham and Trotman, London.
- Joseph, L.N., Field, S.A., Wilcox, C. & Possingham, H.P. (2006) Presence-absence versus abundance data for monitoring threatened species. *Conservation Biology*, **20**, 1679–1687.
- Lennon, J.J., Kunin, W.E., Hartley, S. & Gaston, K.J. (2007) Species distribution patterns, diversity scaling and testing for fractals in southern African birds. *Scaling Biodiversity* (eds D. Storch, P.A. Marquet & J.H. Brown), pp. 51–76. Cambridge University Press, Cambridge.

- MacFadyen, S. (2005) The Kruger National Park CyberTracker Program. *Electronic Ranger Diaries*. http://www.cybertracker.org/KNP_CyberTracker_Article.pdf.
- MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G. & Franklin, A.B. (2003) Estimating site occupancy, colonization and local extinction probabilities when a species is detected imperfectly. *Ecology*, **84**, 2200–2207.
- MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248–2255.
- MacKenzie, D.I., Nichols, J.D., Sutton, N., Kawanishi, K. & Bailey, L.L. (2005) Improving inferences in population studies of rare species that are detected imperfectly. *Ecology*, **86**, 1101–1113.
- Magurran, A.E. & Henderson, P.A. (2003) Explaining the excess of rare species in natural species abundance distributions. *Nature*, **422**, 714–716.
- Mason, T.J. & French, K. (2008) Impacts of a woody invader vary in different vegetation communities. *Diversity and Distributions*, **14**, 829–838.
- McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I. & White, E.P. (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, **10**, 995–1015.
- Nachman, G. (1981) A mathematical model of the functional relationship between density and spatial distribution of a population. *Journal of Animal Ecology*, **50**, 453–460.
- Nachman, G. (1984) Estimates of mean population density and spatial distribution of *Tetranychus urticae* (Acarina: Tetranychidae) and *Phytoseiulus persimilis* (Acarina: Phytoseiidae) based upon the proportion of empty sampling units. *Journal of Applied Ecology*, **21**, 903–913.
- Nichols, J.D., Bailey, L.L., O'Connell, A.F., Talancy, N.W., Grant, E.H.C., Gilbert, A.T., Annand, E.M., Husband, T.P. & Hines, J.E. (2008) Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology*, **45**, 1321–1329.
- Parker, I.M., Simberloff, D., Lonsdale, W.M., Goodell, K., Wonham, M., Kareiva, P.M., Williamson, M.H., von Holle, B., Moyle, P.B., Byers, J.E. & Goldwasser, L. (1999) Impact: toward a framework for understanding the ecological effects of invaders. *Biological Invasions*, **1**, 3–19.
- Pearce, J.L. & Boyce, M.S. (2006) Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*, **43**, 405–412.
- Preston, F.W. (1948) The commonness, and rarity, of species. *Ecology*, **29**, 254–283.
- Pyšek, P., Křivánek, M. & Jarošík, V. (2009) Planting intensity, residence time, and species traits determine invasion success of alien woody species. *Ecology*, **90**, 2734–2744.
- Quinn, G.P. & Keough, M.J. (2002) *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, Cambridge.
- Richardson, D.M. & Whittaker, R.J. (2010) Conservation biogeography – foundations, concepts and challenges. *Diversity and Distributions*, **16**, 313–320.
- Richardson, D.M., Pyšek, P., Rejmánek, M., Barbour, M.G., Panetta, F.D. & West, C.J. (2000) Naturalization and invasion of alien plants: concepts and definitions. *Diversity and Distributions*, **6**, 93–107.
- Rodríguez, A. & Delibes, M. (2002) Internal structure and patterns of contraction in the geographic range of the Iberian lynx. *Ecography*, **25**, 314–328.
- Roura-Pascual, N., Krug, R.M., Richardson, D.M. & Hui, C. (2010) Spatially-explicit sensitivity analysis for conservation management: exploring the influence of decisions in invasive alien plant management. *Diversity and Distributions*, **16**, 426–438.
- Royle, J.A. & Dorazio, R.M. (2008) *Hierarchical Modelling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities*. Academic Press, New York.
- Royle, J.A. & Nichols, J.D. (2003) Estimating abundance from repeated presence absence data or point counts. *Ecology*, **84**, 777–790.
- Royle, J.A., Nichols, J.D. & Kéry, M. (2005) Modelling occurrence and abundance of species when detection is imperfect. *Oikos*, **110**, 353–359.
- Royle, J.A., Kéry, M., Gautier, R. & Schmid, H. (2007) Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecological Monographs*, **77**, 465–481.
- Southwood, T.R.E. (1996) Natural communities: structure and dynamics. *Philosophical Transactions of the Royal Society of London B*, **351**, 1113–1129.
- Strayer, D.L. (1999) Statistical power of presence-absence data to detect population declines. *Conservation Biology*, **13**, 1034–1038.
- Thompson, S.K. (1990) Adaptive cluster sampling. *Journal of the American Statistical Association*, **85**, 1050–1059.
- Thompson, S.K. (2002) *Sampling*, 2nd edn. John Wiley & Sons, New York.
- Thompson, S.K. & Seber, G.F. (1996) *Adaptive Sampling*. John Wiley & Sons, New York.
- Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K. & Possingham, H.P. (2003) Improving precision and reducing bias in biological surveys by estimating false negative error rates in presence-absence data. *Ecological Applications*, **13**, 1790–1801.
- Urquhart, N.S., Paulsen, S.G. & Larsen, D.P. (1998) Monitoring for policy-relevant regional trends over time. *Ecological Applications*, **8**, 246–257.
- Van Kleunen, M. & Richardson, D.M. (2007) Invasion biology and conservation biology: time to join forces to explore the links between species traits and extinction risk and invasiveness. *Progress in Physical Geography*, **31**, 447–450.
- Volkov, I., Banavar, J.R., He, F., Hubbell, S.P. & Maritan, A. (2005) Density dependence explains tree species abundance and diversity in tropical forests. *Nature*, **438**, 658–661.
- Warren, M., McGeoch, M.A. & Chown, S.L. (2003) Predicting abundance from occupancy: a test for an aggregated insect assemblage. *Journal of Animal Ecology*, **72**, 468–477.
- Wilcove, D.S., Rothstein, D., Jason, D., Phillips, A. & Losos, E. (1998) Quantifying threats to imperilled species in the United States. *BioScience*, **48**, 607–615.
- Wilson, R.J., Thomas, C.D., Fox, R., Roy, D.B. & Kunin, W.E. (2004) Spatial patterns in species distributions reveal biodiversity change. *Nature*, **432**, 393–396.
- Wilson, J.R.U., Richardson, R.M., Rouget, M., Procheş, S., Amis, M.A., Henderson, L. & Thuiller, W. (2007) Residence time and potential range: crucial considerations in modelling plant invasions. *Diversity and Distributions*, **13**, 11–22.
- World Conservation Union (2001) *IUCN red list categories and criteria: version 3.1*. IUCN Species Survival Commission, Cambridge.
- Wright, D.H. (1991) Correlations between incidence and abundance are expected by chance. *Journal of Biogeography*, **1**, 463–466.
- Yoccoz, N.G., Nichols, J.D. & Boulinier, T. (2001) Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution*, **16**, 446–453.
- Zhou, S. & Griffiths, S.P. (2007) Estimating abundance from detection-nondetection data for randomly distributed or aggregated elusive populations. *Ecography*, **30**, 537–549.

Received 22 September 2010; accepted 4 January 2011

Handling Editor: Jennifer Finn

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Assessing the accuracy of model prediction under different sampling efforts and schemes, using artificially derived species.

Fig. S1. An illustration of the Bayesian solution to the pseudo-absence dilemma (eqn 2; $M = 100$; $x = 0$).

Fig. S2. The spatial structure of the number of invasive plants in Kruger National Park, as predicted from the Bayesian model. The order of species of the abundance maps is according to Table 1.

Fig. S3. The relationship between sampling effort and the accuracy in abundance estimation.

Fig. S4. The relationship between sampling effort and the accuracy in Moran's I index.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.